Problem Set 2

Denits Panova, Thomas Vicente ,Felix Gutmann

15 May 2016

Problem 1

a)

a.1.)

We start by defining the utility functions for A and B as follows:

$$U_i(A) := \alpha + d_i^A a$$
$$U_i(B) := \beta + d_i^B b,$$

where α, $\beta \geq 0$ and $d_i^* a$ is the number of neighbhours that have chosen A or B.

Using both we derive the best reply. In doing so, start with defining the following inequality (1):

$$U_i(A) > U_i(B)$$

In the utility equations, α and β represent how much choice A is preffered than choice B, having no other node choosing either of the options. They represent the endogenous, independent of the network, utility of respectively choosing A and B.
Proceed by sustaining the definition of the utility functions in the last inequality:

$$\alpha + d_i^A a > \beta + d_i^B b$$

Define d_i to be the sum of d_i^A and d_i^B and rephrase it slightly:

$$d_i := d_i^A + d_i^B$$
$$d_i^B = d_i - d_i^A$$

By substituting the last identity in (1) we can rexpress it in the following way:

$$\alpha + a d_i^A > \beta + b(d_i - d_i^A)$$
$$d_i^A(a+b) \geq \beta - \alpha + b d_i$$
$$d_i^A \geq \frac{\beta - \alpha + b d_i}{b + a},$$

Conclusion:

$\implies$ The best response is to choose A **iff** the former inequality is satisfied.

We repeat this for B. As inequality we get:

$$d_i^B \geq \frac{\alpha - \beta + a d_i}{b + a},$$

a.2.)

The Nash-Equilibrium (NE) occurs when both inequalities are satisfied simultaneously. A NE equilibrium will also be if we always choose B, no matter how many friends have chose A. Such case happens if the endogenous effect β is extremely bigger than α or when b (how much we value the friend who have chosen B< is extremely high.

a.3.)

We can not explain with a p-cluster, because the best response is based on d_i and therefore is not constant. If α and β are zero then we can represent Nash-Equilibria as p-clusters.

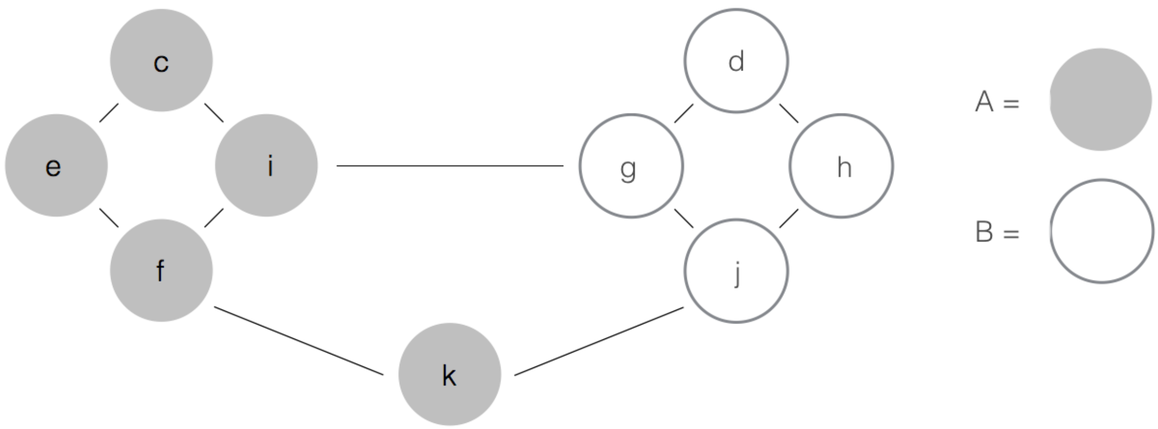
b)

Set $\alpha = \beta = 0$. Each nodes starts with B. Let $t = q = \frac{1}{2}$ for switching to A.

b.1.)

Let $S = \{e - f\}$, where S is the starting set (the two nodes initially adopt behaviour A). In figure 1, we can see which nodes will switch to A.

Figure 1



b.2.)

Let $q = \frac{2}{5}$, so p is going to be$1 - q = \frac{3}{5}$. The p-cluster is $C_p = \{g, h, d, j\}$, with density $D = \frac{2}{3}$, which is bigger than p. See figure 1.

c)

Figure 2: Nash-Equiblibrium 1

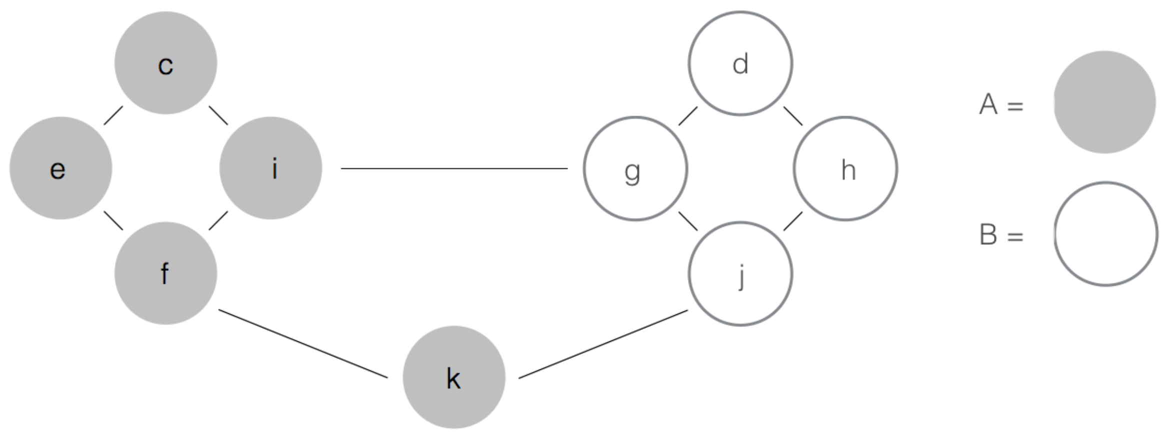


Figure 3: Nash-Equiblibrium 2

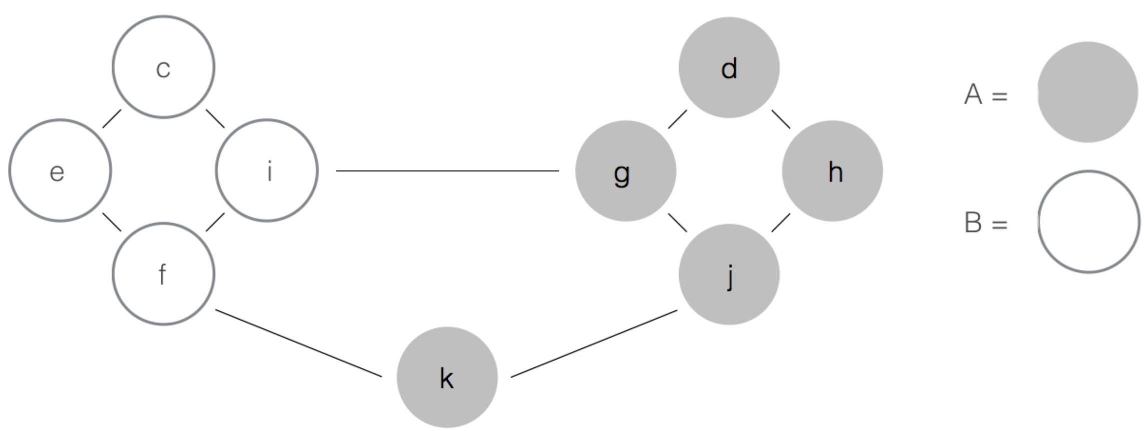


Figure 4: Nash-Equiblibrium 3

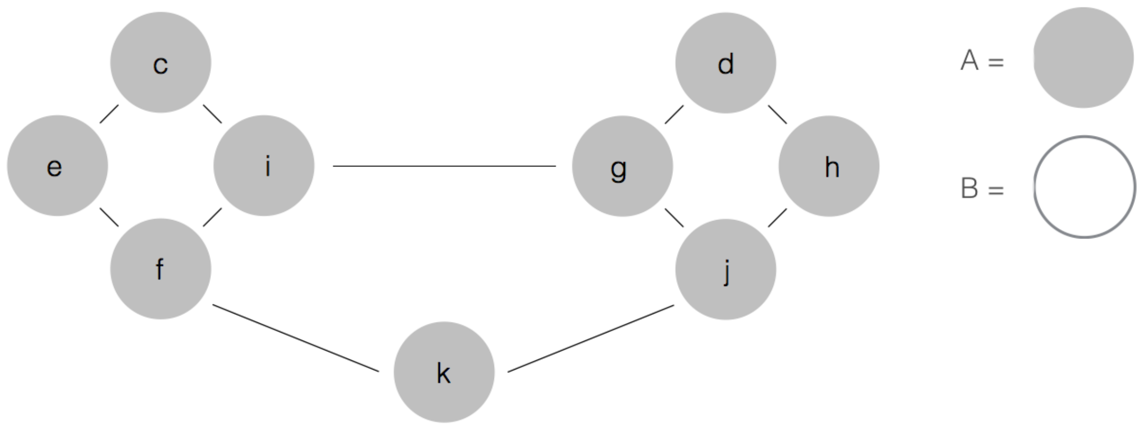


Figure 5: Nash-Equiblibrium 4

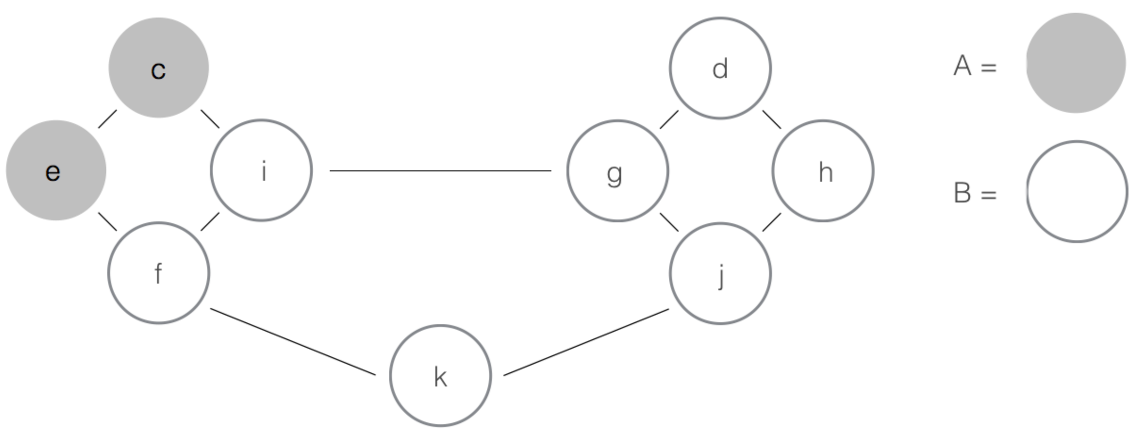


Figure 6: Nash-Equiblibrium 5

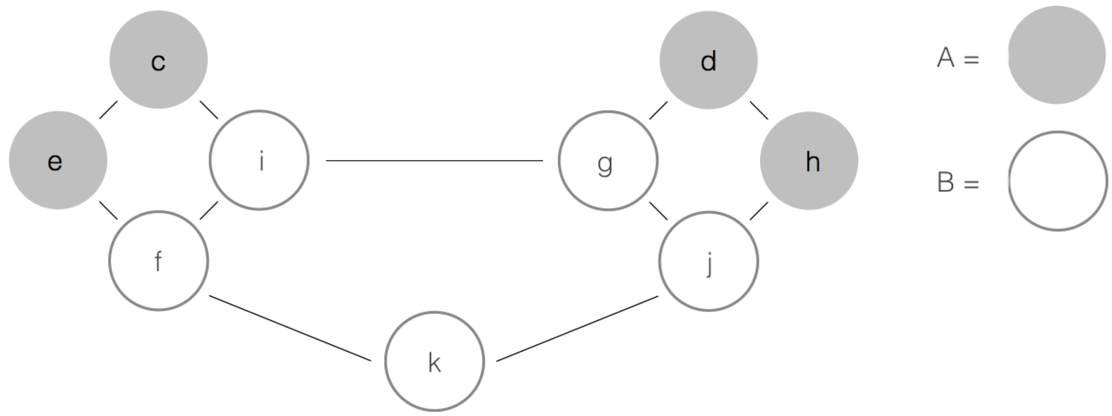
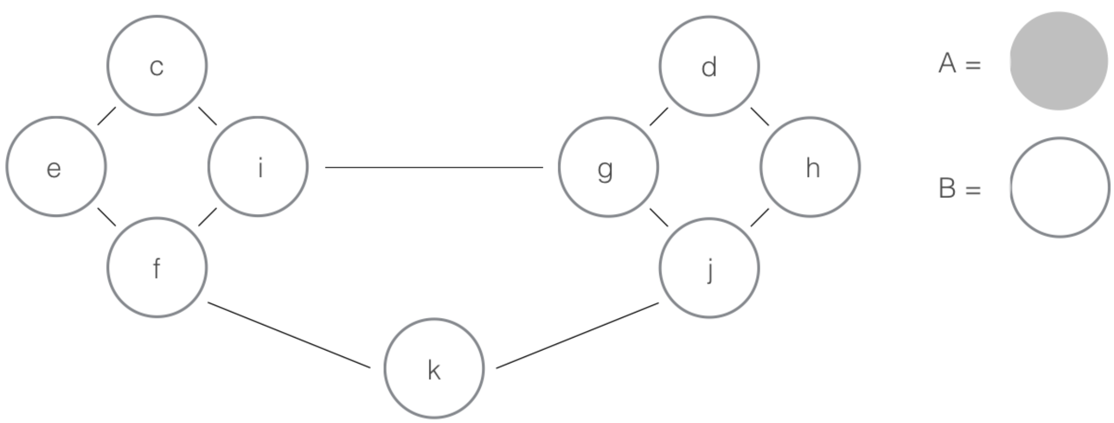


Figure 7: Nash-Equiblibrium 6



d)

We can't have the same Nash-Equilibria. The bigger the difference between α and β is, the more unlikely it is to sustain the initially appointed equilibria. Whichever of both is bigger will point to choosing more frequently option A or B.

Problem 2

2.1)

2.1.a)

We will find the answer within an iterative process:

Step 1:

- $h^{(1)}_{A_i} = 2h^{(1)}_{C_i} = \frac{1}{n} = \frac{1}{8}$
- $a^{(1)}_{B_i} = \frac{3}{8}$
- $a^{(1)}_{D} = \frac{5}{8}$

Step 2:

- $h^{(2)}_{A_i} = \frac{9}{8}$
- $a^{(2)}_{B_i} = \frac{27}{8}$
- $h^{(2)}_{C_i} = \frac{5}{8}$
- $a^{(2)}_{D} = \frac{25}{8}$

Conclusion:

From this procedure we find:

$$a^{(2)}_{B_i} > a^{(2)}_{D}$$

So, even from the second iteration there is a difference between the authority meassures of the two different types of authority nodes.

2.1.b)

We define the formulas for the two sub-graphs separatly. The **first sub-graph** consists of A's and B's. Following that, let $b :=$ number of B's and $a :=$ number of A's and $c =$ constant $= \frac{1}{8}$. For

For $k = 1$:

- $h^{(1)} = a^0 b^0 c = c$
- $a^{(1)} = ab^0 c$

For $k = 2$:

- $h^{(2)} = a^{(1)} b = bab^0 c = b^{(k-1)} a^{(k-1)} c$
- $a^{(2)} = abab^0 c = a^{(k)} b^{(k-1)} c$

For $k = 3$:

- $h^{(3)} = a^{(2)} b = a^{(k-1)} b^{(k-2)} bc = a^{(k-1)} b^{(k-1)} bc$
- $a^{(3)} = aa^{(k-1)} b^{(k-1)} c = a^{(k)} b^{(k-1)} c$

From this procedure we define for the general case:

- $h_{A_i}^{(k)} := a^{(k-1)}b^{(k-1)}bc$
- $a_{B_i}^{(k)} := a^{(k)}b^{(k-1)}c$

We repeat this iterative procedure for the for the **second sub-graph**:

For $k = 1$**:**

- $h^{(1)} = c$
- $a^{(1)} = 5c$

For $k = 2$**:**

- $h^{(2)} = 5c = 5^{(k-1)}c$
- $a^{(2)} = 5 * 5c = 5^k c$

Also for the second graph we define for the general case:

- $h_{C_i}^{(k)} = 5^{(k-1)}c$
- $a_D^{(k)} = 5^k c$

2.1.c)

Firstly, let the normalizing constant be defined as follows:

$$\begin{aligned} const_h &= 3[a^{k-1}b^{k-1}c] + 5 * 5^{k-1}c \\ &= 3 * 3^{2k-2}c + 5^k c \\ &= (3^{2k-1} + 5^k)c \end{aligned}$$

Following that the normalized version of the hub meassures are:

$$\begin{aligned} \widehat{h_{A_i}} &= \frac{h_{A_i}}{const_h} \propto \frac{1}{3}\frac{9^k}{(3^{2k-1} + 5^k)} \\ \widehat{h_{C_i}} &= \frac{5^k}{3^{2k-1}k + 5^k} \end{aligned}$$

The other authority nomalizing constant is define as:

$$\begin{aligned} const_a &= 3a^K b^{k-1}c + 5^k c \\ &= 3^k c + 5^k c \\ \widehat{a_{B_i}} &= \frac{3^{2k-1}}{3^K b^{k-1} + 5^k} \propto \frac{1}{3}\frac{9^k}{3^k + 5^k} \\ \widehat{a_D} &= \frac{5^k}{3^k + 5^k} \end{aligned}$$

Conclusion: As k goes to infitiy $\widehat{a_{B_i}}$ will go to $\frac{1}{3}$ and $\widehat{a_D}$ will converge to 0, assuming that 5^k is sufficinetly small compared to 3^{2k-1}. The same is equivalent for $\widehat{h_{A_i}}$ than $\widehat{h_{C_i}}$. Therefore, the hub vector h is equal to $h = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ and the authority vector is $a = (0, 0, 0, 0, 0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, 0, 0)$, corresponding equivalently for the hub-measure of A's and the authority meassure for the B's.
So when pages have multiple reinforcement, meaning hubs point to more than one page, then in the world web the authorities will be sufficiently big and could bee more easily found.
If we increase the number of C nodes then there is a turning point when the hub-measure of the C's and the authority measure of D will converge to $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. This change of convergence will occur whenever the C nodes are more than nine.

2.2)

The result will be that the authority and the hub-measures will coincide.

$$\begin{aligned}\boldsymbol{h}^{(k)} &= (\boldsymbol{G}\boldsymbol{G}^T)^k \boldsymbol{h}^{(1)} = \boldsymbol{G}^{2k}\boldsymbol{h}^{(1)} \\ \boldsymbol{a}^{(k)} &= (\boldsymbol{G}^T\boldsymbol{G})^{k-1}\boldsymbol{h}^{(1)} = \boldsymbol{G}^{2(k-1)}\boldsymbol{G}\boldsymbol{h}^{(1)} = \boldsymbol{G}^{2k-1}\boldsymbol{h}^{(1)} \approx \boldsymbol{G}^{2k}\boldsymbol{h}^{(1)},\end{aligned}$$

when k goes to infinity.

2.3)

To get convegergence in the HIT algorithm we should normalize at the last step.

$$\begin{aligned}\boldsymbol{h}^{(k)} &= (\boldsymbol{G}\boldsymbol{G}^T)^k \boldsymbol{h}^{(1)} \\ \frac{\boldsymbol{h}^{(k)}}{\omega^{(k)}} &= (\boldsymbol{G}\boldsymbol{G}^T)^k \frac{\boldsymbol{h}^{(1)}}{\omega^{(k)}}\end{aligned}$$

where $\omega^{(k)}$ is the normalizing constant at period k.
As k goes to infinity we converge to $\boldsymbol{h}^*$, but its direction shouldn't change when it is multiplied by $\boldsymbol{G}\boldsymbol{G}^T$.
$\implies \quad \boldsymbol{G}\boldsymbol{G}^T\boldsymbol{h}^* = k\boldsymbol{h}^*$
Therefore $\boldsymbol{h}^*$ is an eigenvector for $\boldsymbol{G}\boldsymbol{G}^T$ and ω is its corresponding eigenvalue.
By power iteration law it follows that ω konverges to the maximum eigenvalue if k goes to infinity.
For any undirected bipartite graph the adjacency matrix is the following:

$$\boldsymbol{G} = \begin{pmatrix} 0 & \boldsymbol{A} \\ \boldsymbol{A}^T & 0 \end{pmatrix},$$

where $\boldsymbol{A}$ is a symmetric matrix. Also $\boldsymbol{G}$ is symmetric.

$$\begin{aligned}\boldsymbol{G}\boldsymbol{G}^T &= \begin{pmatrix} 0 & \boldsymbol{A} \\ \boldsymbol{A}^T & 0 \end{pmatrix}\begin{pmatrix} 0 & \boldsymbol{A}^T \\ \boldsymbol{A} & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & \boldsymbol{A}^2 \\ \boldsymbol{A}^2 & 0 \end{pmatrix}\end{aligned}$$

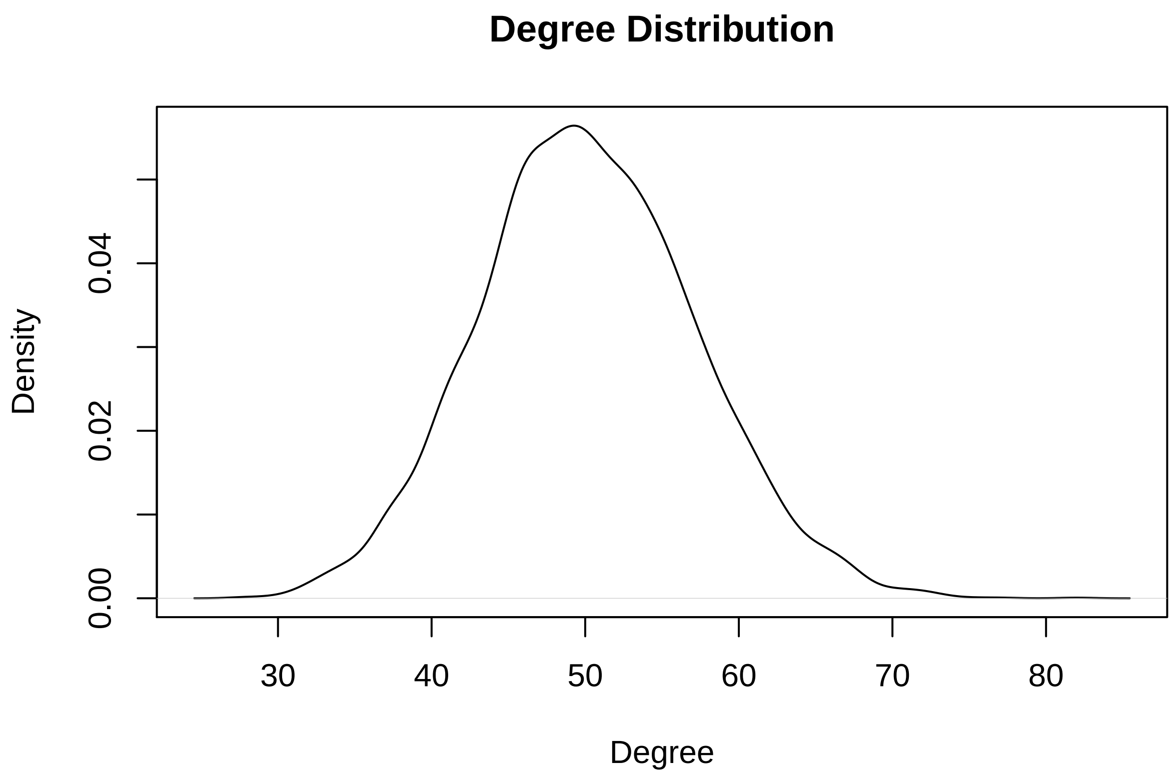
This shows all paths of length two. If we assume that self loops exist then $\boldsymbol{G}\boldsymbol{G}^T$ is actually equal to the projection matrix of the bipartite graph. Thus, $\boldsymbol{G}\boldsymbol{G}^T\boldsymbol{h} = \omega\boldsymbol{h}^*$ is actually the eigenvector centrality measure of the projection matrix, which by definition is

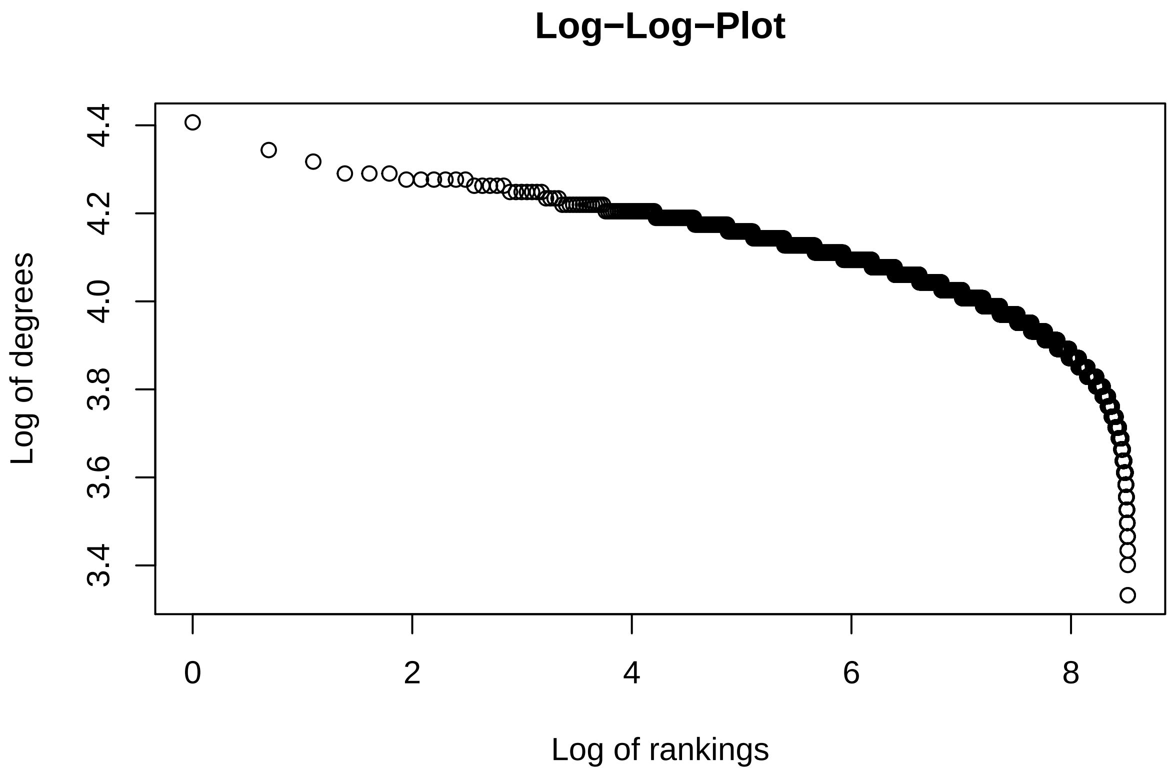
$$\boldsymbol{c} = \frac{1}{\lambda_{max}}\boldsymbol{G}^2\boldsymbol{c}$$

, where $\boldsymbol{c}$ is the centrality vector.

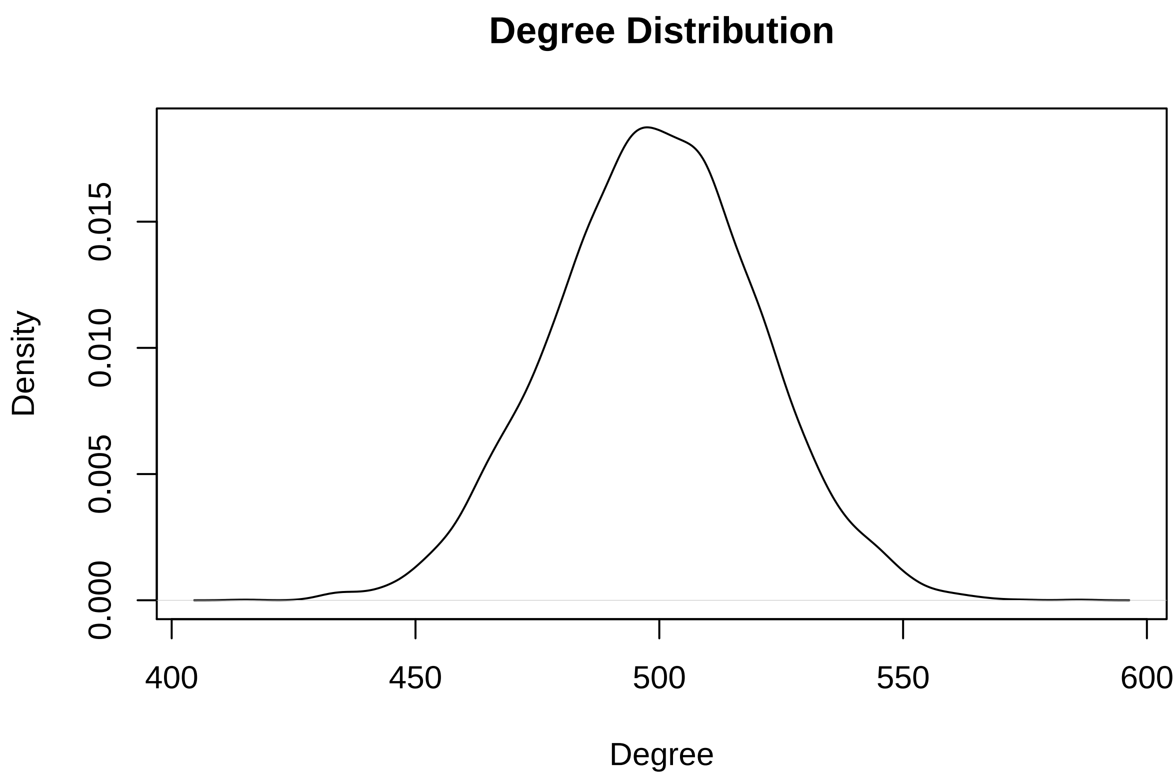
Problem 3

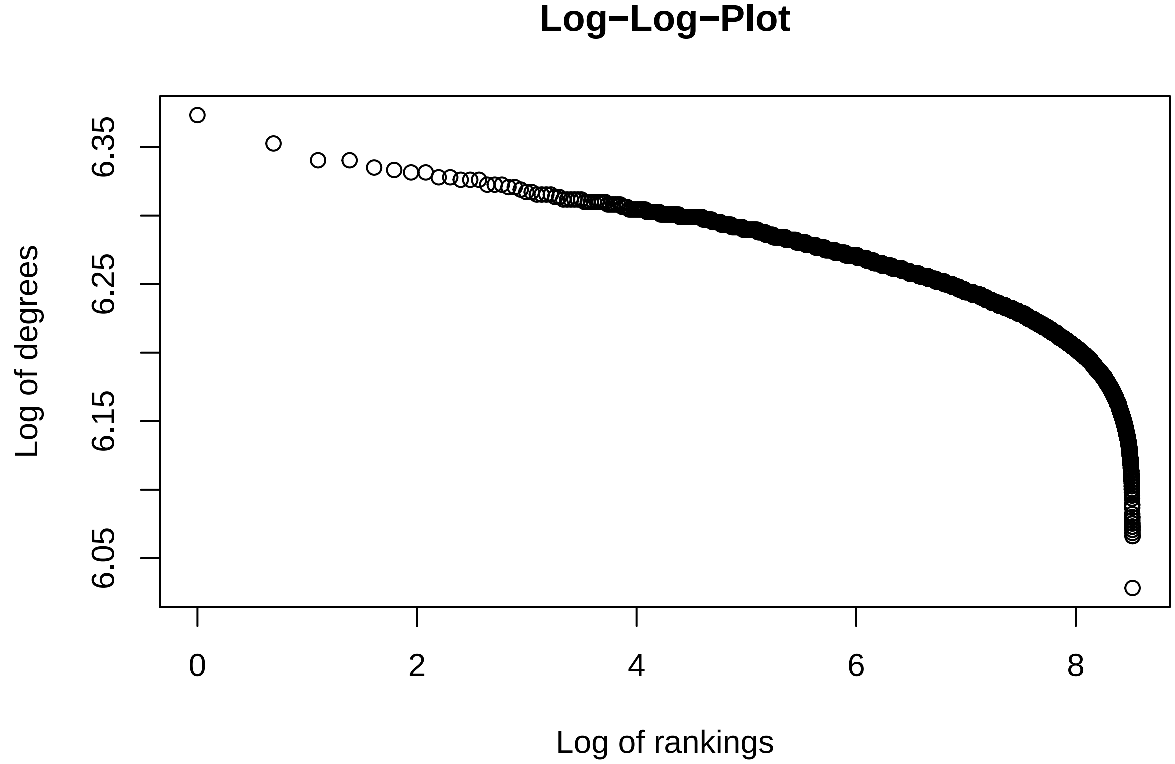
```
-----------------------------------------------------------------------------------------------
For probability = 0.01:
-----------------------------------------------------------------------------------------------
    - average distance = 2.587771
    - diameter = 3
    - the asbolute difference between the average distance and its approximation is 0.410476
    - the asbolute difference between the diamter and its approximation is 0.410476
```





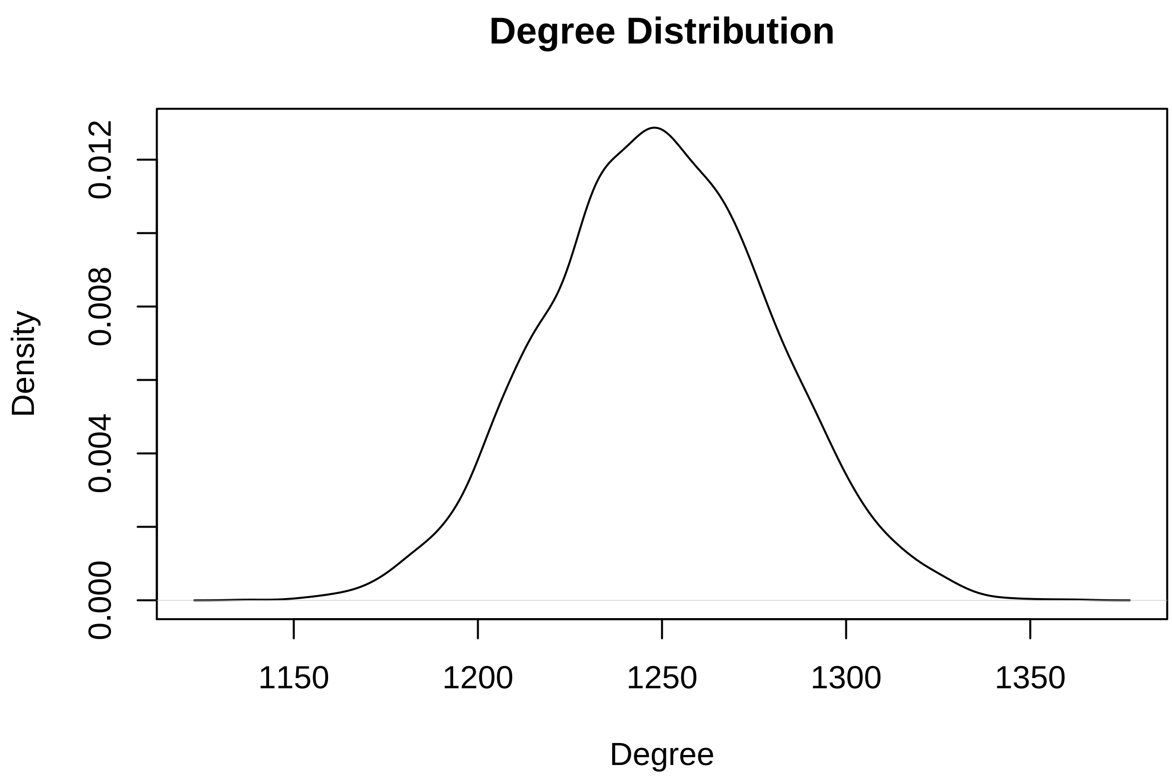

```
    - number of components = 1
    - The numbers of components are: 1
    ...that the largest component captures more than 80 percent of the nodes.


-----------------------------------------------------------------------------------------------
For probability = 0.1:
-----------------------------------------------------------------------------------------------
    - average distance = 1.900136
    - diameter = 2
    - the asbolute difference between the average distance and its approximation is 0.5295802
    - the asbolute difference between the diamter and its approximation is 0.5295802
```

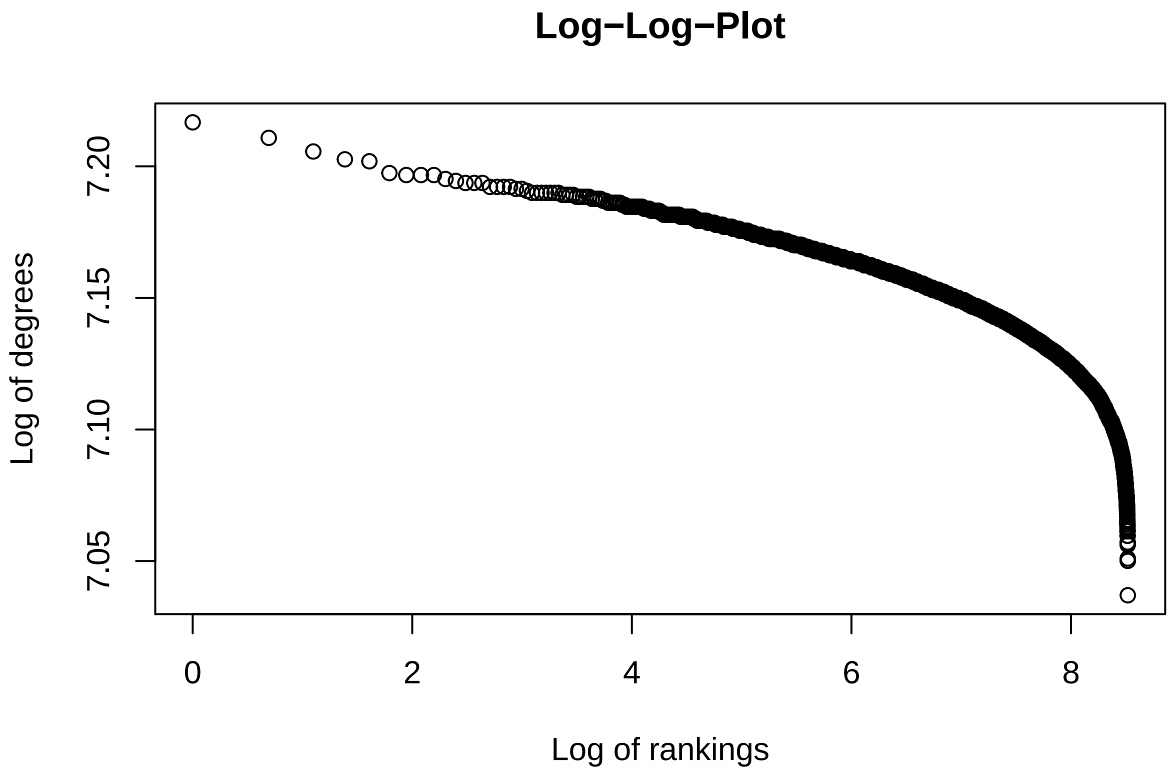




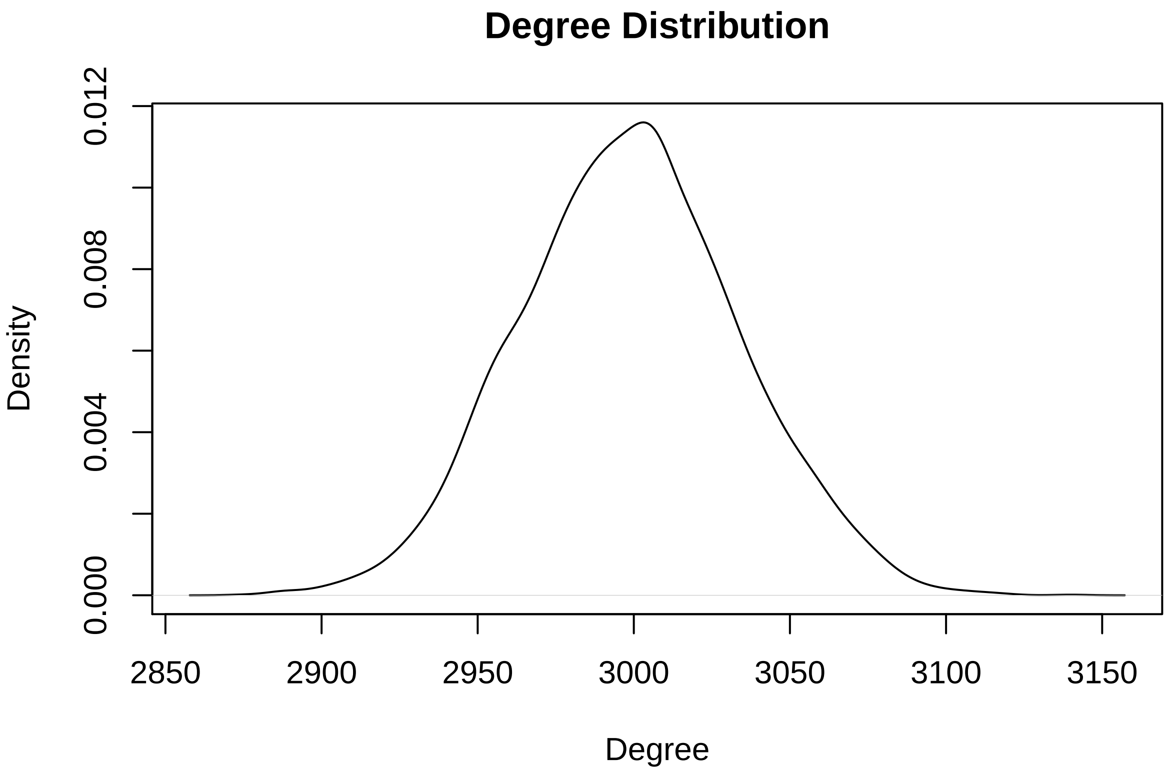

```
- number of components = 1
- The numbers of components are: 1
...that the largest component captures more than 80 percent of the nodes.
```

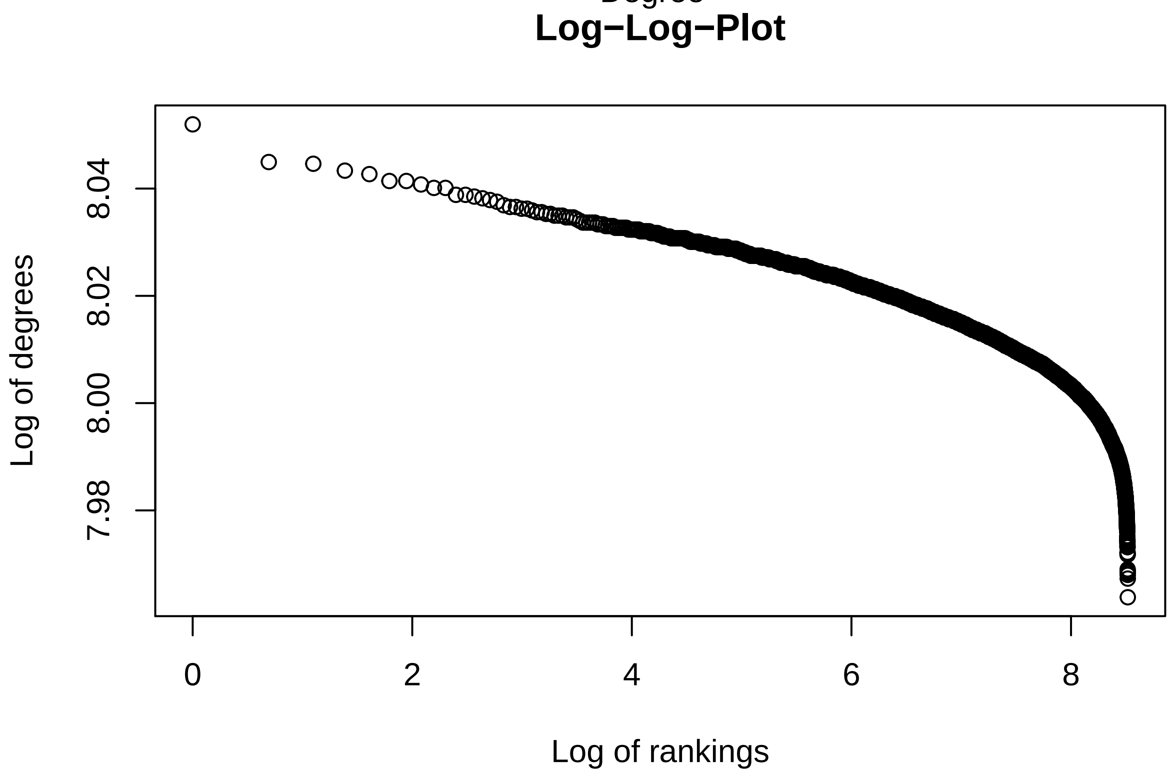
```
--------------------------------------------------------------------------------------------------
For probability = 0.25:
--------------------------------------------------------------------------------------------------
    - average distance = 1.750139
    - diameter = 2
    - the asbolute difference between the average distance and its approximation is 0.555699
    - the asbolute difference between the diamter and its approximation is 0.555699
```






```
    - number of components = 1
    - The numbers of components are: 1
    ...that the largest component captures more than 80 percent of the nodes.


-----------------------------------------------------------------------------------------------
For probability = 0.6:
-----------------------------------------------------------------------------------------------
    - average distance = 1.400138
    - diameter = 2
    - the asbolute difference between the average distance and its approximation is 0.336309
    - the asbolute difference between the diamter and its approximation is 0.336309
```



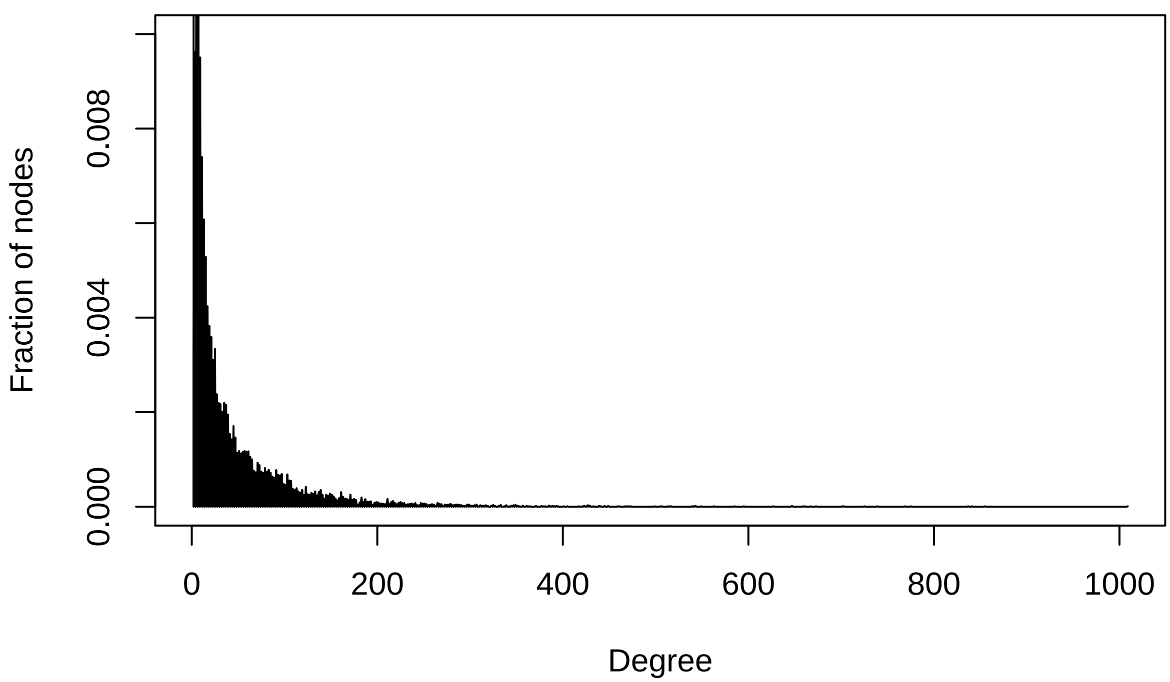



```
- number of components = 1
- The numbers of components are: 1
...that the largest component captures more than 80 percent of the nodes.
```

From the Log-Log-Plot we can see that it looks like a powerlaw distribution.

Problem 4

```
Degree Centrality for the top 10 nodes:
 53213 35290 38109 62821 93504 92790 21718 1086 89732 111161

Eigen Centrality for the top 10 nodes:
 53213 101361 101495 101585 101446 101262 101520 101370 101558 101610

Page Rank for the top 10 nodes:
 105680 81761 54526 92790 50163 22653 23289 39808 92043 36907

HITS for the top 10 nodes:
 53213 37479 124529 84122 127886 10976 90794 91586 95036 75947

Degree Distribution
```



- number of components = 114797

- we see in the following table:

1	2	3	4	5	6	7	8	9	10
114508	140	84	36	14	3	3	3	1	2

12	18	17903
1	1	1

, that the largest component captures 13 percent of the nodes.

But this cluster is still much bigger than the second one.

There are 4054323 adjacent triangles.